

←




Ads by Google


[Stop seeing this ad](#) [Why this ad? ⓘ](#)


IOP Conference Series: Materials Science and Engineering

31


H Index

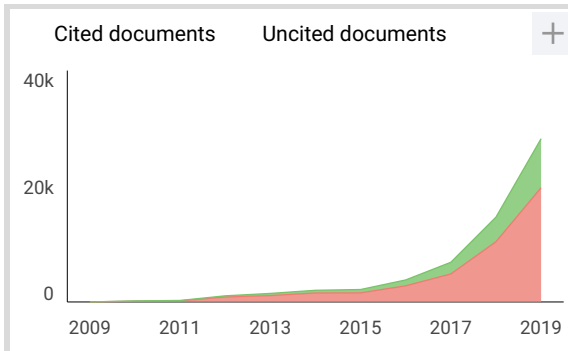
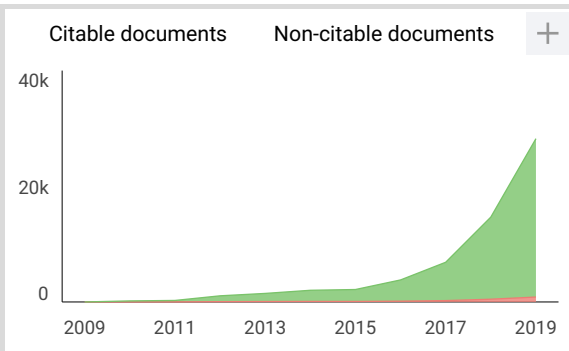
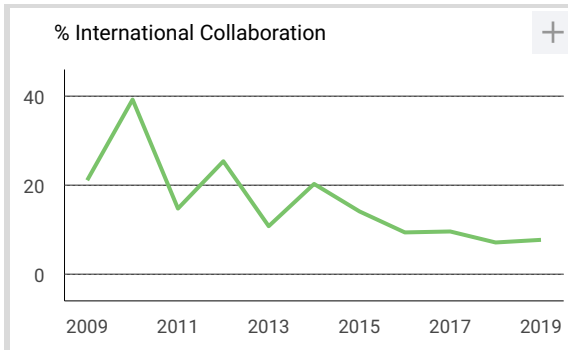
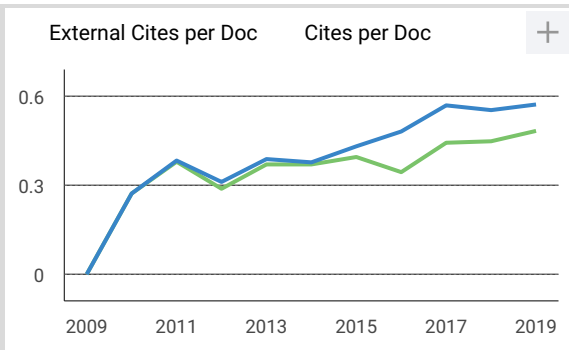
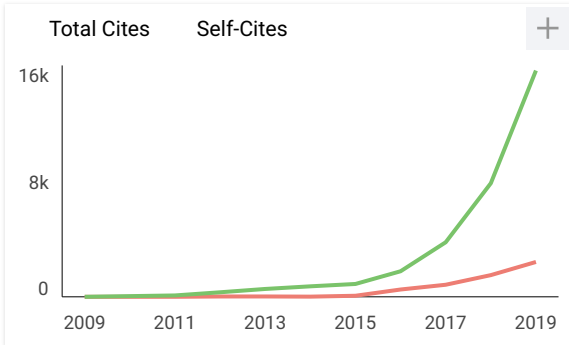
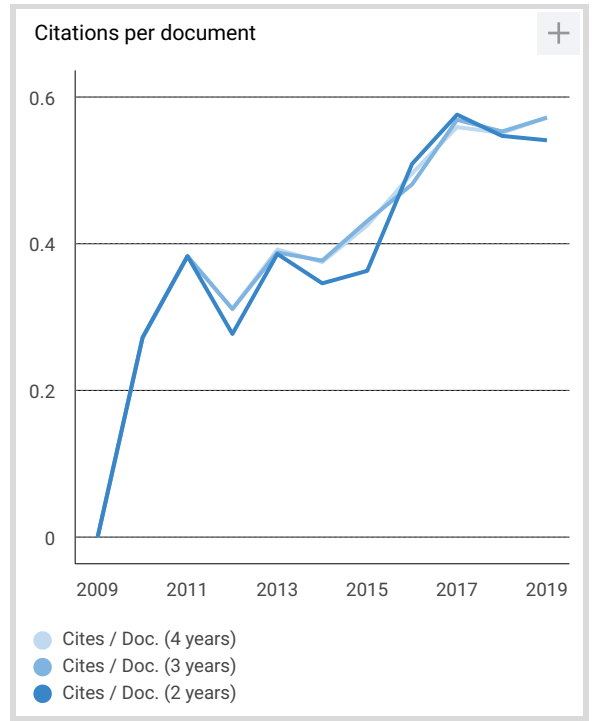
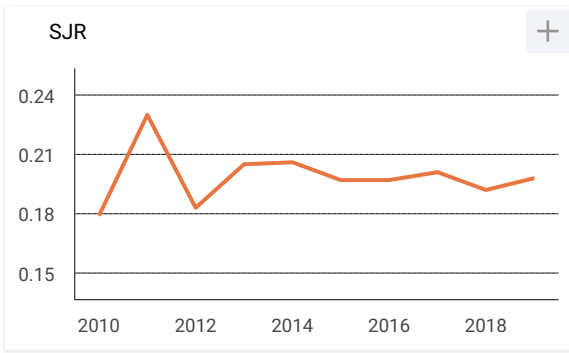
Country	United Kingdom -  SJR Ranking of United Kingdom
Subject Area and Category	Engineering Engineering (miscellaneous) Materials Science Materials Science (miscellaneous)
Publisher	IOP Publishing Ltd.
Publication type	Conferences and Proceedings
ISSN	17578981, 1757899X
Coverage	2009-2020
Scope	The open access IOP Conference Series provides a fast, versatile and cost-effective proceedings publication service for your conference. Key publishing subject areas include: physics, materials science, environmental science, bioscience, engineering, computational science and mathematics.
	Homepage How to publish in this journal Contact
	Join the conversation about this journal


**CV
Templates**


**CV
Examples**


**CV
Builder**


**CV
Help & Tips**



IOP Conference Series: Materials Science and...

← Show this widget in your own website

Not yet assigned quartile

Just copy the code below and paste within your html code:

SJR 2019 **0.2**

powered by scimagojr.com

```
<a href="https://www.scimagojr.com" style="border: 1px solid gray; padding: 2px 5px; display: inline-block;"></a>
```

PAPER • OPEN ACCESS

Vowel Recognition Based on Face Images Using Fisher Linear Discriminant Analysis

To cite this article: Lina and Desi Arisandi 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **852** 012130

View the [article online](#) for updates and enhancements.

Vowel Recognition Based on Face Images Using Fisher Linear Discriminant Analysis

Lina¹, Desi Arisandi^{1*}

¹Faculty of Information Technology, Tarumanagara University, Jakarta, Indonesia

lina@fti.untar.ac.id, *desia@fti.untar.ac.id

Abstract. Speech and voice recognition has a wide range of uses across industries, including embedded devices such as in smartphones, dictation and assistance applications, smart cars, and others. The input for a speech recognition system could be in the form of audio signals or visual images. This paper presents a vowel recognition system, as parts of a speech recognition system, from face images using Fisher Linear Discriminant Analysis (FLDA) method. Images of human faces are used as input for the system. The vowel recognition process consists of the Canny edge detection stage for ROI extraction, the FLDA method for feature extraction, and the Euclidean distance calculation for vowel classification. The output of the system is a written vowel character. The experimental results showed that the average success rate for the vowel recognition was 66%, with vowel “i” and “e” achieved 100% recognition accuracies.

1. Introduction

Nowadays interaction between computers and its users are highly evolved where the devices could read and take instructions from its users. One area of pattern recognition that supports communications between man and machine is speech recognition. A speech recognition system could receive an input in the form of audio signals or visual images. Various techniques have been proposed to obtain a robust speech recognition system from speech utterances, such as the Hidden Markov Model (HMM) [1], the Structured Maximum a Posteriori (SMAP) method [2], the Maximum Likelihood Linear Regression (MLLR) method [3], the Minimum Mean Squared Error Estimation (MMSEE) method [4], and other statistical approaches. Meanwhile, several researches discussed speech recognition systems from visual features, mostly through lips synchronizations. Zhang [5] proposed mouth localization through hue features for developing a speech recognition system, while Wang [6] introduced a lip contour extraction from colour images. Several neural network approaches have also been used in many of the developed speech recognition systems, mainly focused in lips reading [7-8]. In this paper, the research focuses on a simplified version of speech recognition, namely, vowel classification. The system is expected to classify the Indonesian language spoken vowel, such as “a”, “i”, “u”, “e”, and “o” based on face images of the speakers. The system starts by applying the Canny method for edge detection. The Canny edge detection will separate the lip-regions from the facial images. The Fisher Linear Discriminant Analysis (FLDA) method is then applied to extract and determine the lip-features. Finally, the Euclidean distance is used as a classifier. The output of the proposed system is a written vowel character. This article is organized as follows. Section 2 discusses the vowel recognition system based on FLDA method. Section 3 presents the experimental setup and its results. Finally, Section 4 brings the conclusion.



2. The Vowel Recognition System

Generally, a pattern-matching approach involves two essential steps: the pattern training and pattern comparison (testing). In the pattern training stage, several lip segmented images are input to the system and the FLDA method is applied to extract the lip features. In the pattern testing stage, the system captures a face image. The captured image will then be processed using the Canny edge detection method in order to localize the lips areas. Next, the FLDA method is applied to extract the detected lip pattern. Finally, a direct comparison is made between the unknown speech-vowel with each possible pattern learned in the training stage in order to determine the identity of the unknown according to the goodness of match of the patterns. In this step, the Euclidean distance calculation is used as a classifier. The system shows the output in the form of a written text. Figure 1 describes the overall architecture of the proposed vowel recognition system.

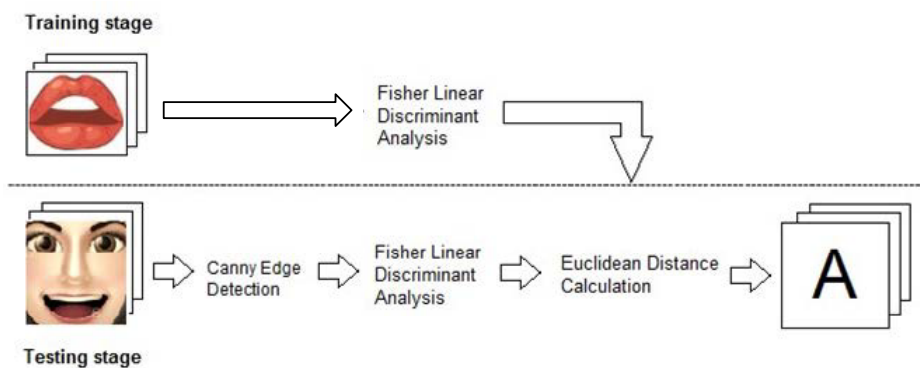


Figure 1. The overall process of the proposed vowel recognition system

2.1. Canny edge detection

Canny edge detection is an algorithm with several steps that can detect edges with noise suppressed at the same time. Canny algorithm uses 2D Gauss function derivative for handling the original image [9]. First, a preprocessing step which changes RGB images to grayscale images is applied to each input. The Canny edge detection starts with a smoothing process to reduce noise in the image with Gaussian Filter. The equation for applying the Gaussian filter is as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where G is the Gaussian filter and σ is the standard deviation. The smoothness level of the image is highly dependent on the selection of σ . The second step of the Canny edge detection is to compute the

image gradient using any of the gradient operators such as Sobel, Prewitt, Roberts, etc. The equation for calculating the image gradient (G) and the edge direction (θ) are as follows:

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (2)$$

$$\theta = \arctan\left(\left|\frac{G_y}{G_x}\right|\right) \quad (3)$$

The third step of the Canny edge detection is to perform non-maxima suppression which aims to remove a pixel from an edge candidate if the pixel is not a local maximum value. This iterative process is performed until a thin edge is obtained. The next step of the Canny edge detection is the thresholding step. In this step, a threshold is defined to convert a grayscale image into a binary image while maintaining all edge elements in its place. The last step of the Canny edge detection is hysteresis which is useful to eliminate disjointed lines in an object.

2.2. Fisher Linear Discriminant Analysis (FLDA)

FLDA is a feature extraction method with a combination of mathematical and statistical calculations that impose separate statistical properties for each object. The purpose of the FLDA method is to search for linear projections to maximize the between-class covariance matrix while minimizing the within-class covariance matrix, so that members in the class are more dispersed and be able to improve the success of recognition [10]. The calculation of the FLDA method is as follows:

$$\sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad 4$$

$$N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad i=1$$

where S_W is the within-class covariance matrix, S_B is the between-class covariance matrix, c is the number of classes, N_i is the number of images in class- i , μ_i is the average image in class- i and μ is the average of the pixel values of the overall images.

The eigenvalue (λ) and the eigenvector (V) of the covariance matrix are then calculated based on the within-class covariance matrix (S_W) and the between-class covariance matrix (S_B):

$$\det(VS_BV^T) \quad 6$$

$$\det(VS_WV^T)$$

$$S_BV = \lambda S_WV \quad 7$$

Once the eigenvector is obtained, the f_x value of the FLDA feature can be calculated using this equation:

$$f_x = \sum_{i=1}^k (x_i - \mu)^T \times V \quad (8)$$

2.3. Euclidean Distance

Euclidean distance is a formula to calculate distance of a feature (x_i) to its neighbours (y_i) as shown in Eq. 9. The unknown feature is classified in the same class as a feature that has a minimum distance.

$$d_{x,y} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (9)$$

3. Experimental Results

This section describes the experiments conducted for the proposed vowel recognition system using the FLDA method. We developed our own database, called the FTI-Untar vowel-image database, which consists of a total of 50 lip images with 10 images for each vowel class, used for training. There were 5 vowel-classes in Indonesian language, i.e. “a”, “i”, “u”, “e”, and “o”. For testing, we used 100 face images consist of persons who were pronouncing the vowels. Figure 2 shows the image sample of each process in the testing stage. First, the original image was input to the system. Next, the image was converted to a grayscale image for Canny edge detection. The result of the Canny edge detection method was the ROI image of the original image. The process was continued with a feature extraction process by applying the FLDA method. Once the recognition process was completely done using the Euclidean distance calculation, the system generated the output character based on the recognized vowel as shown in column 4 in Figure 2. The screen capture of the developed vowel recognition application is presented in Figure 3.

Furthermore, Table 1 shows the recognition accuracies for each vowel class in the experiments. The highest recognition accuracies were achieved by vowel classes “i” and “e” with 100%. However, the recognition results dropped significantly on vowel class “o” with 55% recognition accuracy, while class “u” and “a” obtained only 40% and 35% recognition accuracy, respectively. Several conditions that potentially triggered false recognition in the proposed system were variations in lighting during the face capturing process and lip movements of some people that were not changing much when pronouncing one vowel to another.

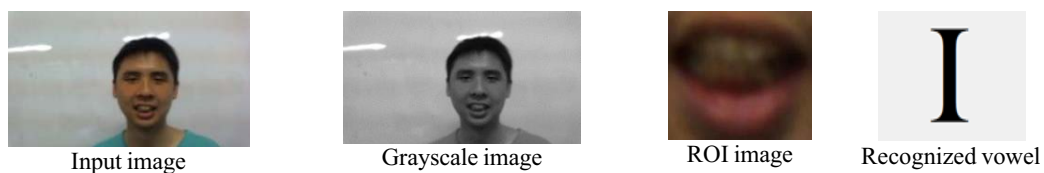


Figure 2. The image sample of each process in testing stage

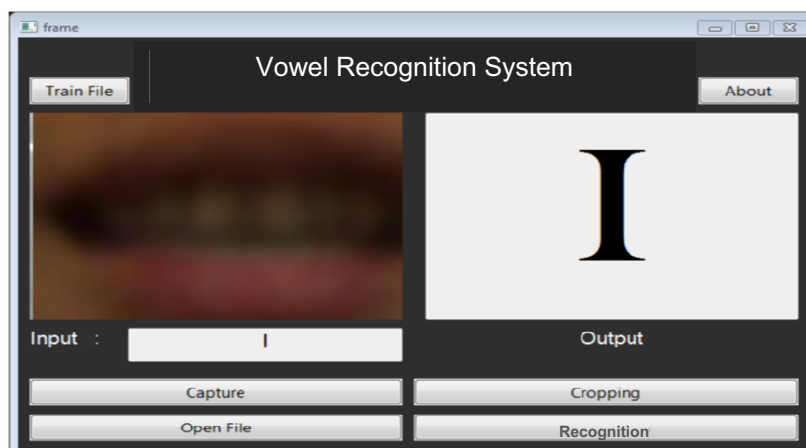




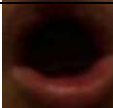


Figure 3. The screen capture of the developed vowel recognition application

Table 1. Recognition accuracies for each vowel class

Class	Image sample	Accuracy (%)
A		35
I		100
U		40
E		100
O		55
Average		66

4. Conclusion

This paper has discussed a vowel recognition system based on face images using the Fisher Linear Discriminant Analysis (FLDA) method. Experimental results showed that the combination of the FLDA method as a feature extractor and the Euclidean distance method as a classifier could give 100% accuracy for some vowel classes such as “i” and “e”. However, the system needs more improvement on the recognition of the other vowel classes, i.e. “a”, “u”, and “o”. The average recognition results for the vowel recognition system was 66%. Future works will be focused on modifying the classifier to improve the system’s robustness and accuracy.

5. References

- [1] Chien JT 1999 *IEEE Trans. on Speech and Audio Proc.* **7** 656
- [2] Shinoda K and Lee CH 2001 *IEEE Trans. on Speech and Audio Proc.* **9** 276
- [3] Leggetter CJ and Woodland PC 1995 *Computer Speech and Language* **9** 171
- [4] Indrebo KM 2008 *IEEE Trans. on Audio, Speech and Language Proc.* **16** 1654
- [5] Zhang X, Mersereau RM 2000 *Proc. of Int. Conf. on Image Proc. (Vancouver BC, Canada)* pp 226-229
- [6] Wang SL, Lau WH, Leung SH 2004 *Pattern Recognition* **37** 2375
- [7] Matthews L 2002 *IEEE Trans. on Pattern Analysis and Machine Intel.* **24** 198
- [8] Sadeghi VS, Yaghmaie K 2006 *Int. J. of Computer Science and Network Security* **6** 154
- [9] Hosseini MM, Gharahbagh AA, Ghofrani S 2010 *Proc. of Int. Conf. on Knowledge-based and Intelligent Inf. and Eng. Sys. (Cardiff, UK)* pp 331-339
- [10] Xiao feng Zhang, Yu Zhang, Ran Zheng, 2011 *Procedia of Engineering* **15** 1335
Rachmad A, Devie R, Anamisa N 2017 *Advanced Science Letters* **23** 12344